Genotype Imputation in Nelore Cattle

R. Carvalheiro[1], J. Sölkner[2], H. Neves[1], Y.T. Utsunomiya[1], A.M. Pérez O'Brien[2], S. Boison[2], M. da Silva[3], C.P. VanTassell[4], T.S. Sonstegard[4], J. McEwan[5], F.S. Schenkel[6] , J.F. Garcia[1]

[1]Universidade Estadual Paulista, São Paulo - Brasil, [2]University of Natural Resources and Life Sciences, Vienna - Austria, [3]EMBRAPA, Juiz de Fora - Brasil, [4]USDA- ARS, Beltsville - USA, [5]AgResearch, Invermay - New Zealand, [6]University of Guelph, Guelph - Canada

Objective

Test imputation efficiency in Nelore using different SNP densities and software

- Population vs. Pedigree based

Conclusions

1. FImpute outperformed BEAGLE in imputation accuracies
2. Pedigree information increased accuracy just for 6k
3. Relatives in reference panel increased accuracy (BEAGLE)
4. LD between markers was important for imputation accuracy

Table 1. Mean percentage ± SD of correctly called genotypes (PERC) and correlation between imputed and observed genotype (CORR) for different scenarios of customised SNP panels

SNP panel	BEAGLE		FImpute – Pedigree		FImpute – Pedigree free	
	%PERC	CORR	%PERC	CORR	%PERC	CORR
6k	87.7 ± 4.4	89.8 ± 3.9	90.6 ± 4.1	92.6 ± 3.5	89.7 ± 4.1	91.6 ± 3.5
11k	96.1 ± 2.3	97.0 ± 1.9	97.6 ± 1.5	98.2 ± 1.1	97.6 ± 1.6	98.2 ± 1.2
15k_Ev	95.5 ± 2.5	96.4 ± 2.1	97.2 ± 1.8	97.8 ± 1.4	97.2 ± 1.7	97.8 ± 1.3
15k_LD	96.4 ± 2.2	97.1 ± 1.8	97.9 ± 1.4	98.4 ± 1.1	97.8 ± 1.5	98.4 ± 1.1
46k	98.2 ± 1.2	98.6 ± 1.1	99.1 ± 0.7	99.3 ± 0.5	99.1 ± 0.7	99.3 ± 0.5

Figure 1. Effect of genomic relationships between reference and test set on imputation accuracy

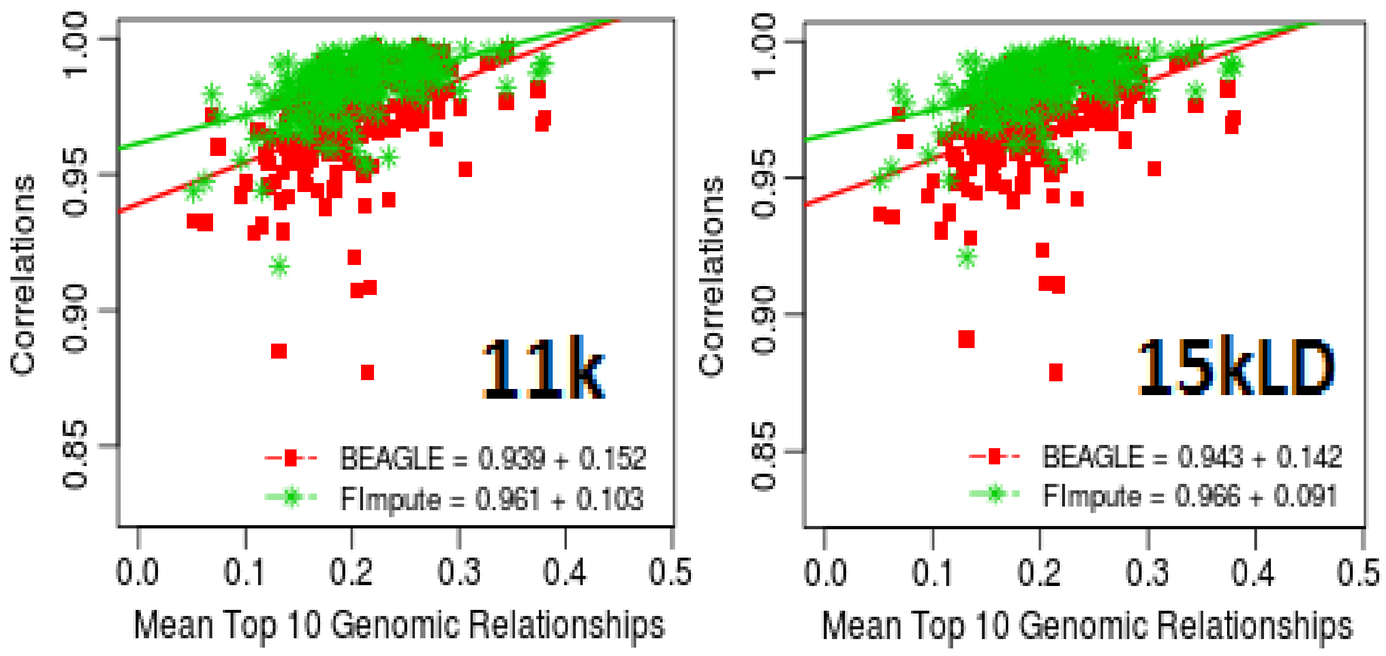
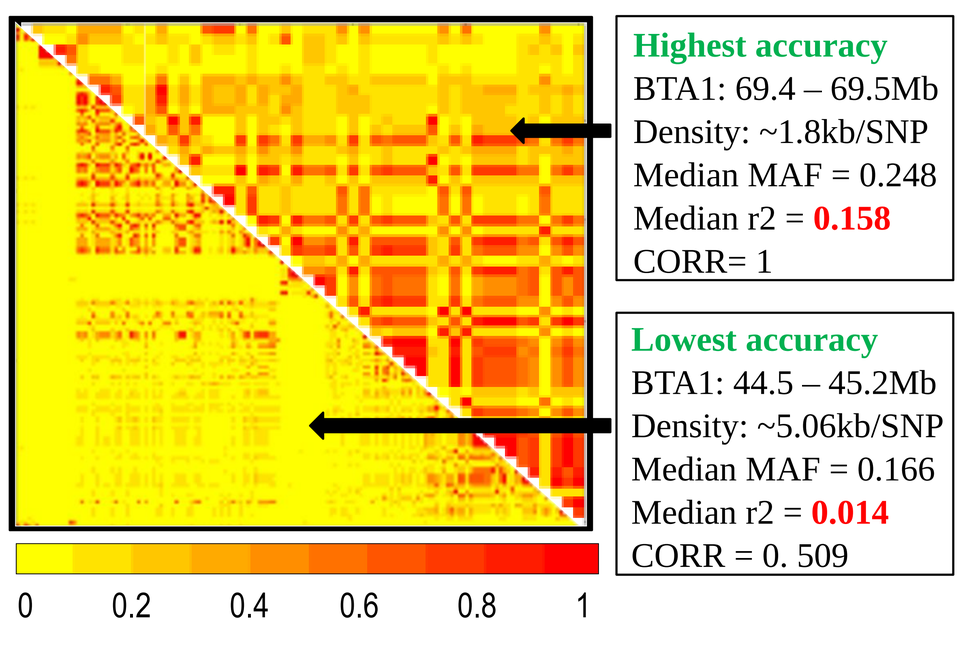


Figure 2. Impact of Linkage Disequilibrium on accuracy



Background

- Genotype imputation reduce costs of breeding programs implementation
- Available low density panels developed using *Bos taurus* breeds
- Need to developed alternative low SNP panels for *Bos indicus* (Nelore breed)

Materials and Methods

- 793 sires used as reference and 202 sires used as testing set
- Illumina Bovine HD (777k SNPs)
- Quality control: MAF > 0.02, call rate > 0.98, GC score > 0.7 ➔ 439,595 SNPs

Scenarios

Illumina 6k, 11k (Illumina 6k + customised*), evenly spaced15k (Ev), 15k customised*(LD), 46k (Illumina 6k + customised*)

***SNPs with the highest MAF and LD within 20 SNP-windows across the genome**

Software

Fimpute, BEAGLE